

SYSTEM AND METHOD FOR PROVIDING NEWS UPDATES

Field of the Invention

The present invention relates generally to the provision of content updates relating to specified subjects. More particularly, the present invention relates to the provision of updates of content appearing on the World Wide Web (hereinafter "Web") relating to specific subjects.

Background

A user of the Web typically gathers news about a subject of particular interest by actively searching the Web for relevant news items. Such a search is very time consuming and typically retrieves both desired news articles and undesired content containing common search terms. Structuring a search so as to avoid unwanted content without severely limiting the breadth of the search is typically difficult if not impossible.

Alternatively, the user may sequentially browse a collection of Web sites known to him or her to be possible sources of news items relating to the subject of interest. However, the user will need to search or browse each Web site separately and will not retrieve news items present only on other Web sites.

What is needed is an automated tool for retrieving timely news articles about a subject of interest, without also

09607879-063000

retrieving unwanted content, from a wide variety of relevant Web sites.

It is therefore an object of the present invention to provide a system and method for providing items, other than advertisements, about a subject of interest to an individual.

Summary of the Invention

00607879 "063000

A system and method are provided for providing news relating to a specified subject to a subscriber, wherein a plurality of Web sites relating to a category to which the specified subject relates is selected, the relevancy of at least one Web page in each selected Web site is determined by scanning for words relating to the specified subject, the content type of at least one Web page in each selected Web site is determined by scanning for words indicating content type, a list of relevant Web pages is compiled based on the determinations of relevancy and content type, and the compiled list is provided to the subscriber.

Brief Description of the Drawings

Figure 1 is a block diagram of a system in accordance with a preferred embodiment of the present invention.

Figure 2 illustrates the flow of data among computers in accordance with a preferred embodiment of the present invention.

Figure 3 is a flow chart of a method in accordance with a preferred embodiment of the present invention.

Detailed Description of the Preferred Embodiments

Referring to figure 1, a preferred embodiment of a system in accordance with the present invention is illustrated. Web server 100 may be a mainframe, minicomputer, microcomputer, or other type of computer (or may be composed of a plurality of computers connected by a network or other means), but is preferably a Windows NT or Unix server, including at least processor 102, such as a Pentium family processor, and memory 104 connected thereto. Memory 104 may be temporary memory, such as random access memory, or permanent storage, such as a hard drive, but is preferably a combination of temporary memory and permanent storage. News Update software 106, stored in memory 104, in a first preferred embodiment is written in Perl and JAVA and uses a regular expression algorithm to descramble universal resource locators (URLs) located in target Web pages and to identify specific elements pertaining to a particular subject. Once these elements are identified, a string parser breaks the target down into tokens and, based on the frequency of each token, categorizes the document accordingly.

Web server 100 is connected by Internet connection 108 to the Internet 110. User terminal 114 is similarly connected to the Internet 110 by Internet connection 112. Internet connections 108 and 112 may be direct connections, such as T-1 lines or indirect connections, such as modem-to-modem

connections over telephone lines, or any other sort of connection to the Internet, and Internet connections 108 and 112 may be of different types.

User terminal 114 may be a dumb terminal or a computer, such as a mainframe computer, a minicomputer, a desktop or laptop microcomputer, a personal digital assistant, or a smartphone. User terminal 114 is most typically a personal computer with a Pentium family processor running a Windows operating system. User terminal 114 is connected to display 116, which may be a cathode ray tube or liquid crystal display monitor, although in some cases user terminal 114 may be integrated with display 116 in a single unit. User terminal 114 is also connected to at least one of keyboard 118 and pointing device 120, which may be a mouse or trackball. In some cases keyboard 118 and pointing device 120 may also be integrated into user terminal 114.

Referring to figure 2, in the preferred embodiments of the present invention, a user submits a search request, as described below in connection with figure 3, from user terminal 114 to Web server 100, which accesses Web site database 202 and search term database 204 in response to the request. Web site database 202 and search term database 204 may be relational, object oriented, or other custom or commercial off-the-shelf databases, such as Oracle version 8.

Web site database 202 contains entries for each Web site that may be searched in accordance with the method described

below in connection with figure 3 indicating the category or categories to which such Web site pertains. The categories may be categories such as sports, music, politics, fashion, or technology, may be geographically-based categories (e.g., New York, Mid-West, Europe, etc.), may be based on age, gender, ethnicity, religion, vocation, or avocation, or may be based on a combination of any or all of such categories. Moreover, a category may be very general, such as music or sports, or very specific, such as Madonna or Troy Aikman. As described below, this table may be created manually or automatically.

Search term database 204 may contain search terms relating to categories, to particular subjects within categories, or to both. For example, search term database 204 might contain terms useful in searching for articles on sports, terms useful in searching for articles on American football, terms useful for searching for articles on a particular football team, or some combination of such types of terms.

After formulating a search, as described below in connection with figure 3, Web server 100 searches a plurality of Web sites 210a through 210n based on the contents of Web site database 202. Within each of the n Web sites, at least one Web page is searched 212a through 212n, 214a through 214n, or 216a through 216n respectively, using search terms drawn from search term database 204 or from the user (or some combination of such terms). The results are then returned to user terminal 114.

09607879-063000

Referring to figure 3, a method in accordance with a preferred embodiment of the present invention is illustrated. In the preferred embodiments hereunder, as a part of step 300, before the performance of step 300, or after the performance of step 300, in at least the first iteration of the present method with respect to a particular user, the user is prompted to provide a subject about which the user desires to receive news articles (or other non-advertising content) or references to news articles (or other non-advertising content). The user may be so prompted by displaying a message on a Web page requesting that the user specify a subject and providing the user with a text entry box or a drop-down list box for supplying the subject, or the user may be provided with the opportunity to navigate through a pre-indexed Web site with hyperlinks to popular subjects. The user may also be so prompted by an e-mail message prompting the user to reply by e-mail with the desired subject (in the header or body of the message) or by other means. In other embodiments of the present invention, the user may not be prompted to provide a subject at all. For example, in an embodiment directed to the employees of a particular corporation, the employees might automatically receive content relating to the corporation or relating to their job functions. Similarly, members of a professional, recreational, or political organization might automatically receive content relating to the organization or subject matter related to such organization.

The user may also be prompted to select a category into which the subject falls from a predetermined list of categories.

Alternatively, the news update software may categorize the subject automatically from stored subject category combinations (e.g., New York Giants/Sports, C++/Technology, AARP/Senior Citizens) or the administrator of the news update service may categorize the subject manually.

In addition, the user may optionally be prompted to select search terms to be used by the news update software.

Preferably, the user is first presented with a list of search terms that the news update software will use by default and the user is then offered the opportunity to add or delete terms from the list. Alternatively, the news update software can rely on the user to provide all search terms or can automatically use the default terms in each case.

In step 300, a plurality of Web sites relating to the category to which the specified subject relates is selected. This plurality may be selected by accessing Web site database 202 and retrieving a set of Web sites pertaining to the same category to which the specified subject relates. In other embodiments, the user may be prompted to supply, or modify, the list of Web sites, or the list may be determined manually by the news update software administrator. In any event, if the method has previously been performed with respect to a particular user with respect to the same subject, only Web sites containing content that has been altered since the last performance of the method with respect to the user with respect to the same subject are selected.

prior to the performance of any of the steps of the present method. In any event, the content type of a Web page may be determined to be a particular type based on a single occurrence of a word for which the Web page is scanned, based on the occurrence of a predetermined number of such words, or based on the proportion of words suggesting a particular content type to words suggesting another (or all other) content type or types.

In step 306, a list of relevant Web pages is compiled. Preferably, the list consists of all Web pages determined to relate to the selected subject based on the results of step 302 and also determined to be of an appropriate content type (e.g., news or all content other than advertising) based on the results of step 304.

In step 308, this list is provided to the subscriber (or other user in other embodiments of the present invention). The list may be displayed on a Web site, or sent by e-mail, ordinary mail, facsimile, an HTML or XML feed, beeper, cell phone, or other means, but is preferably sent by e-mail to the user. The list includes at least the uniform resource locator (hereinafter the "URL") of each Web page in the list and may also include the date on which each Web page was last modified, the title of the document, and the publication source.

Optionally, feedback may be utilized to improve the accuracy or speed of the present method. For example, Web sites whose Web pages consistently fail to generate any hits in step 302 or consistently are categorized as being of an inappropriate content type (such as advertising) may be omitted from

